**Movie Rating System Using Collaborative Filtering**

* **Introduction**: Collaborative filtering is a mechanism to predict user interests by gathering preferences from many users. For a movie rating system, it predicts how a user would rate an unseen movie.
* **Problem Representation**:
  + r(i,j)r(i,j): Whether user jj has rated movie ii (1 if rated, 0 otherwise).
  + y(i,j)y(i,j): Rating given by user jj to movie ii.
  + mm: Number of movies.
  + nn: Number of features (e.g., genres).
  + x(i)x(i): Feature vector for movie ii.
  + θ(j)θ(j): Parameter vector for user jj.
* **Predicting Movie Ratings**:
  + For each user jj and movie ii, the predicted rating is (θ(j))Tx(i)(θ(j))Tx(i).
  + A linear regression model can be set up for each user.
* **Cost Function**:
  + A regularized linear regression cost function is used to minimize the difference between the predicted ratings and the actual ratings.
  + Parameters for all users are simultaneously adjusted.
* **Collaborative Filtering**:
  + The system can infer movie features (like genres) based on user feedback.
  + Using this method, both the movie features and user preferences can be simultaneously optimized to give the best predictions.
* **Algorithm Steps**:
  + Initialize x(i)x(i) and θ(j)θ(j) to small random values.
  + Use gradient descent (or an advanced optimization algorithm) to minimize the cost function.
  + Predict star ratings using the optimized xx and θθ values.
* **Vectorization**:
  + Use Low Rank Matrix Factorization to simplify calculations. Given matrices XX and ΘΘ, the matrix of all predicted ratings YY is Y=XΘTY=XΘT.
  + The similarity between two movies can be gauged using the distance between their respective feature vectors.
* **Implementation Detail: Mean Normalization**:
  + To handle new users with no movie ratings, mean normalization is introduced.
  + Normalize the ratings by subtracting the mean rating for each movie.
  + Adjust the prediction equation to include the mean normalization term, ensuring that predictions for new users aren't initialized to zero but to the mean rating of each movie.